1. Lensing basics

SHERRY H. SUYU

The field of gravitational lensing has evolved from a theoretical fantasy to a robust astrophysical tool in the past few decades. In this chapter, we introduce the basics of gravitational lensing. We explain what gravitational lensing is in Section 1.1 and briefly recount the history of lensing in Section 1.2. We continue with the basic theory for lensing in Section 1.3 and work through properties of simple lens mass distributions in Section 1.4.

1.1. Introduction

A perhaps familiar example of lensing is the bending of light by optics, such as the glasses that some people wear, or binoculars that some people use for viewing wildlife or events. These two examples of optical lenses are linear in the sense that one sees only a single (perhaps magnified) image of the object of interest. In contrast, the base of a wine glass is a non-linear lens so that through the glass one can see multiple images of the background object. Figure 1.1 is an illustration of this phenomenon. In the top-left panel, we see the original source of light that emanates from the candle. In the top-right panel, we see four images of the source as lensed by the wine glass: one in the lower left, two close pairs on the lower right and one behind the stem of the wine glass. By tilting the base of the wine glass, we change the properties of the optical lens and thus the light paths of the images we see. For example, in the bottom-right panel, there are only two multiple images of the background source. In the case where the stem of a symmetric wine glass is aligned perfectly along of the line of sight to the source, the source is lensed into a ring, as shown in the bottom-left panel.

In gravitational lensing, a massive object takes on the role of the lens, similar to the wine glass in optical lensing. According to Einstein's General Theory of Relativity, mass curves spacetime. Light, taking the shortest path in this curved spacetime, thus 'bends' around massive objects. Anything that has mass (e.g. planets, stars, galaxies, and clusters of galaxies) can thus act as a gravitational lens. In Figure 1.2, we show an example of an elliptical galaxy acting as a gravitational lens: a background galaxy (a.k.a. 'source') is lensed into four images by the foreground elliptical galaxy (a.k.a. 'lens') in the centre. Note that the morphology of the lensed images is similar to that in the top-right panel of Figure 1.1. The ellipticity of the lens galaxy is analogous to the tilt of the base of the wine glass.

There are three regimes of gravitational lensing: strong, weak, and micro. In strong lensing, the source, lens, and observer are closely aligned and the source is lensed into multiple images. In cases where the background source is spatially extended, the images are strongly distorted into rings/arcs. Figure 1.2 is an example of a strong lens. In weak lensing, the source lies further away from the line connecting the observer and the lens. Consequently, there are only weak distortions of singly imaged background sources. The small distortions of the images due to weak lensing are typically not obvious to the eye of the observer since the distortions in the images are smaller than the intrinsic shapes of the sources. In microlensing, the lenses are stars (possibly with planets) either in
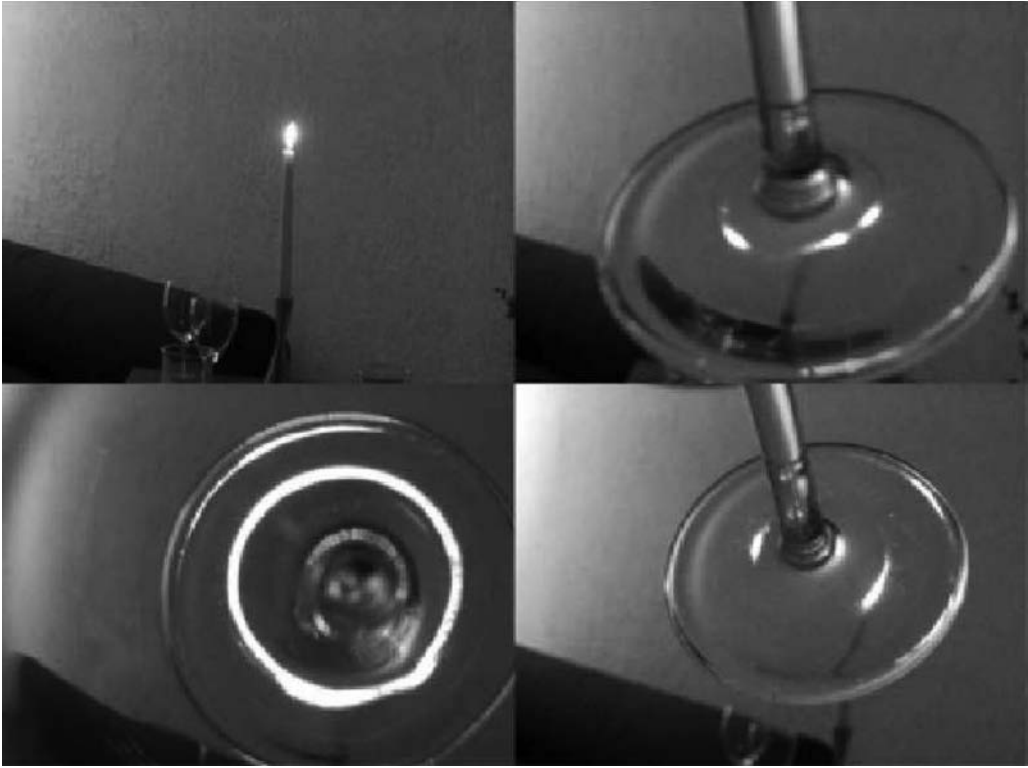
1

Sherry H. Suyu

FIGURE 1.1. Wine glass: an example of a non-linear optical lens. Top left: the source (candle flame) in the absence of any lens. Top right and bottom right: the multiple images of the source as seen through the base of a wine glass. The image configuration and morphology depend on the tilt of the wine glass. Bottom left: the image of the source forms a ring when the stem of the wine glass is aligned along the line of sight to the source. (Image courtesy of Phil Marshall.)
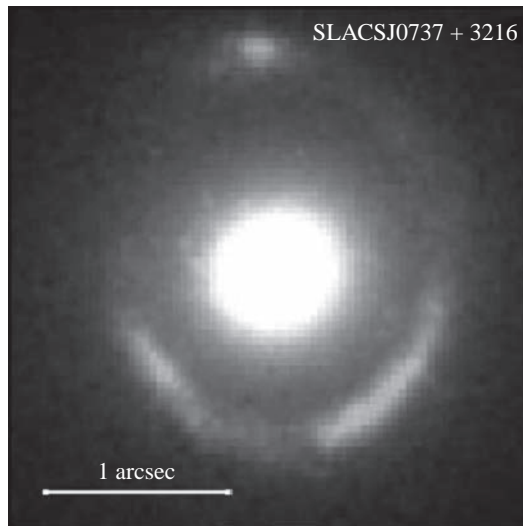

FIGURE 1.2. The gravitational lens SLACSJ0737+3216. A background source galaxy is lensed into four images by the foreground elliptical lens galaxy in the centre. (Data from SLACS and Marshall et al. 2007, *ApJ*, **671**, 1196; image courtesy of Phil Marshall.)

our own Milky Way or in other galaxies. A star, when aligned between the observer and a background source, leads to multiple images of the source; however, the separations between the multiple images are of the order of microarcseconds (hence the name, 'microlensing'). This image splitting on microarcsecond scales is not resolved by telescopes; observers simply see a brightening of the background source that results from the formation of unresolved multiple images as the lens passes between the observer and the source.

1.2. Brief history of lensing

We briefly recount here the history of lensing. For more details, we refer the readers to, for example, Schneider, Kochanek and Wambsganss (2006).

1.2.1 *Deflection of light*

The deflection of light rays by massive objects has been predicted for centuries. According to Newton's theory of gravitation, a test particle moving at velocity v past an object of mass M is deflected by an angle

$$\hat{\alpha} = \frac{2GM}{v^2\xi}, \tag{1.1}$$

where G is the gravitational constant and ξ is the impact parameter. If one treats light as (massive) particles, then the Newtonian deflection angle of light is

$$\hat{\alpha}_{\mathrm{N}} = \frac{2GM}{c^2\xi}, \tag{1.2}$$

where c is the speed of light. This was first speculated by John Mitchell in 1784 and later by Johann von Soldner in 1804.

When Einstein completed his General Theory of Relativity (GR) in 1915, the behaviour of light in a gravitational field could be considered without assuming light as a stream of massive particles. According to GR, the deflection of light by an object of mass M is

$$\hat{\alpha}_{\mathrm{E}} = \frac{4GM}{c^2\xi} = 1.75'' \left(\frac{M}{M_\odot} \right), \tag{1.3}$$

which is twice the Newtonian prediction in equation (1.2).

As seen in equation (1.3), the deflection angle is proportional to the mass of the object and thus is most easily detected near massive objects. In the vicinity of our planet Earth, the most massive object is the Sun, so various expeditions set out in the early 1900s to observe the gravitational deflection of light by the Sun. In 1919, Sir Arthur Eddington and collaborators measured the positions of stars near the Sun (in projection) during a total solar eclipse and confirmed Einstein's predictions for the deflection angles.

1.2.2 *Discovery of a strong lensing system: a double quasar*

It took until several decades later for the discovery of the first strong gravitational lens system. Walsh, Carswell and Weymann (1979) detected a pair of quasars separated by $\sim 6''$. Figure 1.3 shows the optical images of the system, known as QSO 0957+561. The short exposure in the left panel clearly shows the two QSOs marked by A and B, and the longer exposure in the right panel reveals the lens galaxy G1 in between A and B. The confirmation of this as a gravitational lens system comes from the spectra of the two QSOs, which are identical.
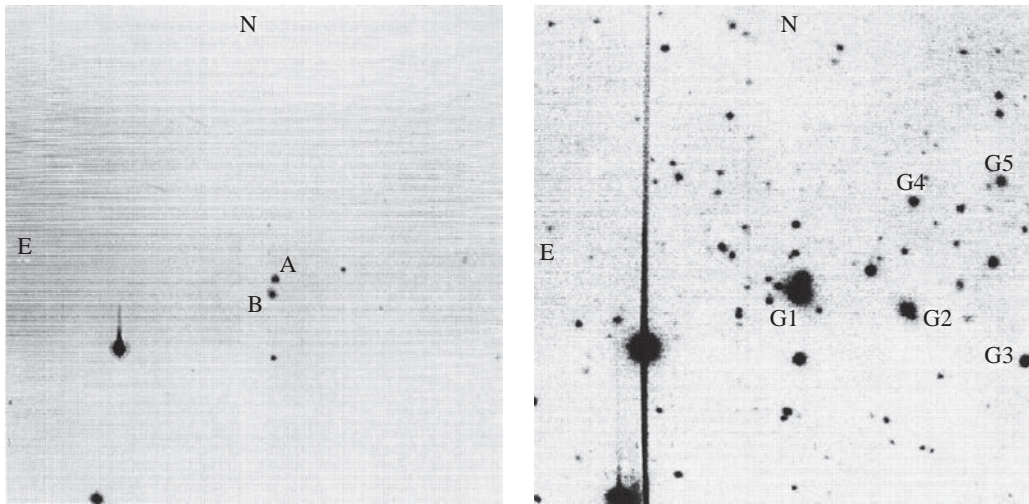
FIGURE 1.3. The first detection of strong gravitational lensing: double QSO 0957+561. Left: short optical exposure, showing the two QSOs, marked A and B, that are separated by 6″. Right: longer optical exposure reveals the lens galaxy G1 between A and B, the extended source of the QSO host galaxy, and the small cluster of galaxies of which G1 is the brightest member. (Reproduced with permission from Young et al. 1980, *ApJ*, **241**, 507.)

1.2.3 *Discovery of giant luminous arcs*

Several years after the detection of the double QSO, two independent teams (Lynds & Petrosian 1986; Soucail et al. 1987) discovered giant luminous arcs around two clusters of galaxies (see Figure 1.4). Each arc is displaced from the centre of the galaxy cluster and curves around it. The nature of these arcs was unclear at the time, and several hypotheses were put forward as possible explanations. All but one (Paczyński 1987) of these hypotheses were rejected as a viable explanation when the redshift of the arc was measured by Soucail et al. (1988) to be much higher than the redshift of the cluster. It was then clear that the arc corresponded to strongly magnified, lensed images of a higher-redshift galaxy. With the advent of the *Hubble Space Telescope* (*HST*), these arcs are now routinely detected in galaxy clusters; in fact, there can even be dozens of strongly lensed background sources in a single galaxy cluster, as in Abell 1689.

1.2.4 *Detection of quasar microlensing*

As we saw in the case of strong lensing, the global mass distribution of a galaxy can lens a background source into multiple images. Figure 1.5 shows an image of the lens system QSO 2237+0305, where the bulge of a spiral galaxy lenses a background QSO into four images. The individual stars of the lens galaxy in front of each QSO image can act as microlenses and change the fluxes of the QSO images independently. Since the stars orbit in the galaxies and there is relative motion between the microlenses, source and observer, the magnifications of each QSO image due to microlensing vary with time. By observing the quasar lens system QSO 2237+0305 at different epochs, the quasar microlensing effect was first detected in this lens system by Irwin et al. (1989).

1.2.5 *Detection of Galactic microlensing*

The stars in our Milky Way act as lenses for other stars or extragalactic sources. The light curve of the background star as it is lensed has a characteristic shape shown in
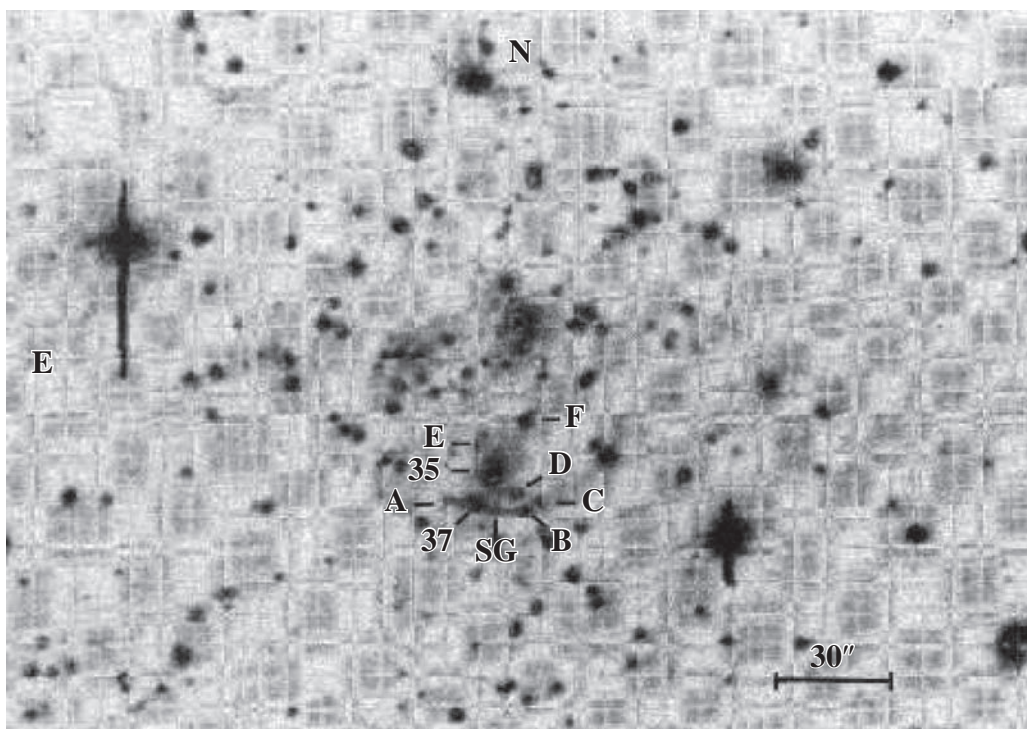
FIGURE 1.4. Giant luminous arc in Abell 370. One of the first detections of an elongated, curved arc close to a luminous elliptical galaxy near the centre of the galaxy cluster. (Reproduced with permission from Soucail et al. 1987, *A&A*, **172**, L14.)
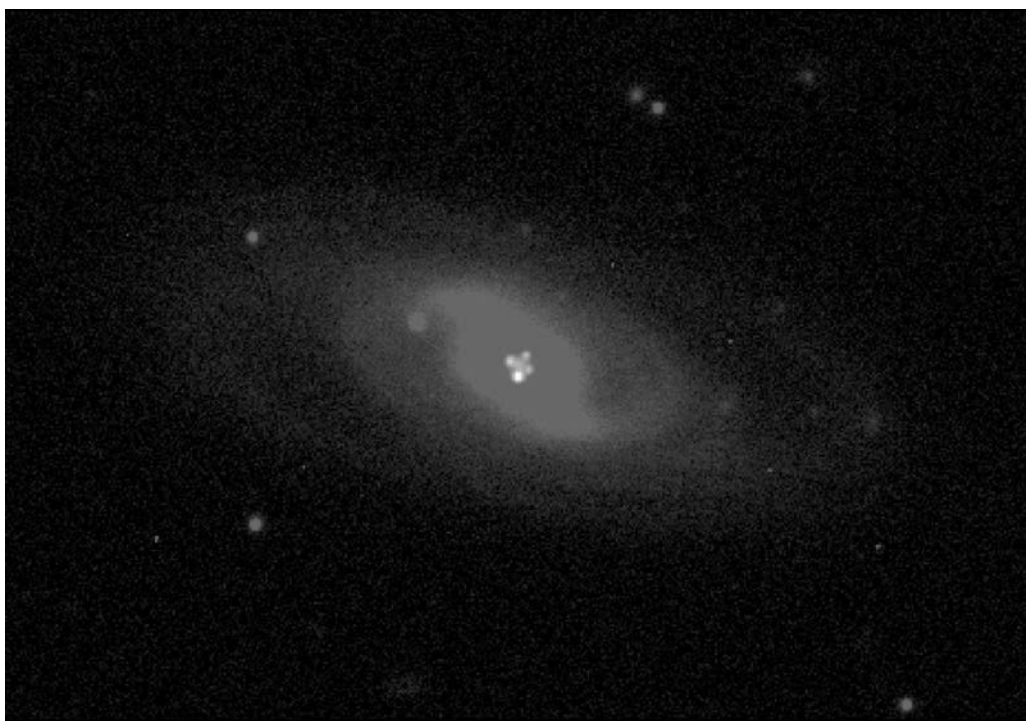

FIGURE 1.5. Near-infrared image of the gravitational lens QSO 2237+0305 where a background quasar is lensed into four images by the bulge of a spiral galaxy at $z = 0.0395$. Microlensing of quasar images was first detected in this lens system by Irwin et al. (1989). (Image courtesy of José Muñoz.)
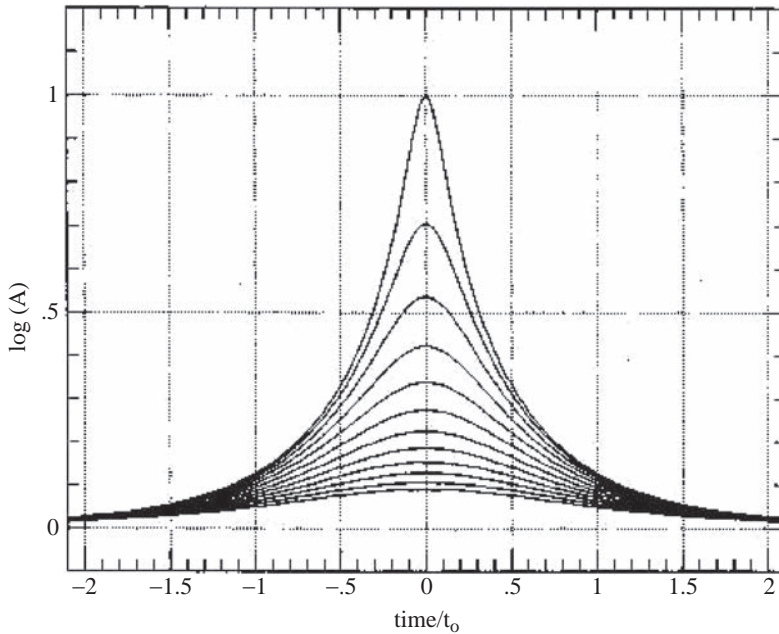
FIGURE 1.6. Microlensing amplification of a background source from the relative motion of the microlens and the source in the plane of the sky. The various curves are for different impact parameters between the background source and the microlens: the smaller the impact parameter, the higher the amplification of the background star. (Reproduced with permission from Paczyński 1986, *ApJ*, **304**, 1.)

Figure 1.6 (Paczyński 1986): the smaller the impact parameter of the lensing star, the higher the magnification of the background star. The probability of such a lensing event is low, with only around 1 in 10^7 stars being lensed at any given time. Nonetheless, by carrying out large monitoring surveys of individual stars in the Large Magellanic Cloud in the early 1990s, two research teams detected Galactic microlensing events (Alcock et al. 1993; Aubourg et al. 1993). Figure 1.7 shows the light curve of a background star that was monitored in the MACHO Project and displayed the characteristic shape shown in Figure 1.6 (Alcock et al. 1993). Since then, Galactic microlensing has become a unique astrophysical probe, particularly in detecting extrasolar planets around stars that act as lenses. The presence of a planet around the foreground star leads to a distinct secondary peak that is embedded within the characteristic microlensing amplification light curve.

1.2.6 *Detection of weak lensing*

As mentioned at the beginning of the chapter, weak lensing occurs when the background sources are sufficiently distant from the lens mass distribution in projection such that the sources have only single lensed images whose shapes are weakly distorted, or 'sheared'. Since the distortion is typically much smaller than the intrinsic shape of the background galaxies, one needs to measure and average over the shapes of a local ensemble of galaxies to infer the amount of shear produced by the foreground mass distributions. This weak gravitational lensing was first detected in two galaxy clusters by Tyson, Wenk and Valdes (1990), and even weaker lensing around individual galaxies was discovered by Brainerd, Blandford and Smail (1996).
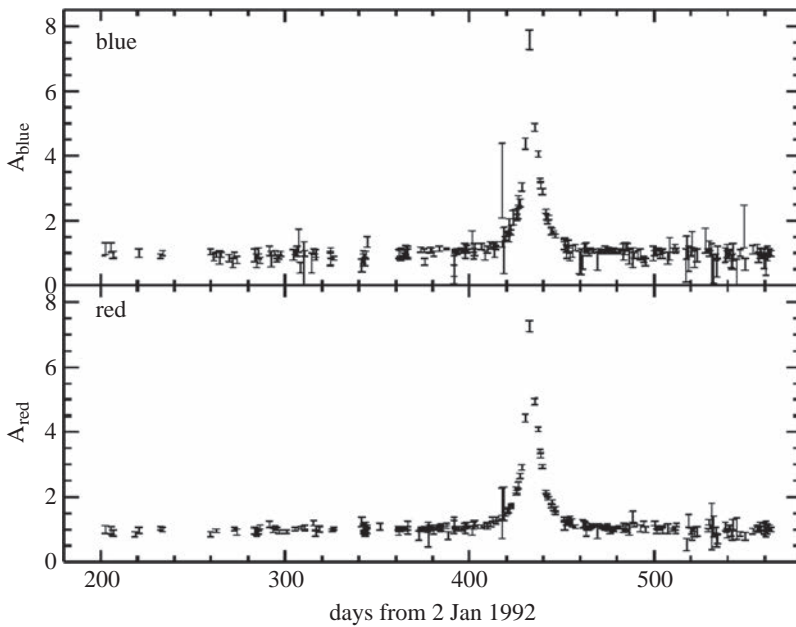
FIGURE 1.7. Galactic microlensing event detected in the MACHO Project. The microlensing amplification of a background source was detected in both the blue and red passbands. (Reproduced with permission from Alcock et al. 1993, *Nature*, **365**, 621.)

1.2.7 *Lensing in the new millennium*

Radio, infrared, optical imaging, and spectroscopic surveys in recent years have led to an explosion in the discovery of lens systems. These surveys include, for example, the Cosmic Lens All-Sky Survey (CLASS: Myers et al. 2003; Browne et al. 2003), the Herschel Astrophysical Terahertz Large Area Survey (Negrello et al. 2010), the South Pole Telescope and the Atacama Large Millimeter/submillimeter Array observations (Vieira et al. 2013; Hezaveh et al. 2013), the SDSS Quasar Lens Search (SQLS: Oguri et al. 2006; Inada et al. 2012), the Strong-Lensing Legacy Survey (SL2S: Limousin et al. 2009; Gavazzi et al. 2012), the HST Archive Galaxy-scale Gravitational Lens Search HAGGLeS: Marshall et al. 2009), the Cosmic Evolution Survey (COSMOS: Faure et al. 2008; Jackson 2008), the Major UKIDSS-SDSS Cosmic Lens Survey (MUSCLES: Jackson et al. 2012), the Sloan Lens ACS Survey (SLACS: Bolton et al. 2006, 2008), the Sloan WFC Edge-on Late-type Lens Survey (SWELLS: Treu et al. 2011), and the BOSS Emission-Line Lens Survey (BELLS: Brownstein et al. 2012).

There are now hundreds of known lens systems, permitting unprecedented studies of mass structures and cosmology. In particular, gravitational lensing is a unique probe of dark matter since it is sensitive to the lens mass distribution, independently of the light profile or dynamical state of the lens. Lensing has yielded interesting results on the properties of dark matter haloes of galaxies, groups of galaxies, and clusters of galaxies (e.g. Mandelbaum et al. 2006; Natarajan et al. 2009; Auger et al. 2010; Suyu & Halkola 2010; Newman et al. 2013; Okabe et al. 2013), and on the detection of dark matter substructure (e.g. Dalal & Kochanek 2002; Fadely & Keeton 2012; Vegetti et al. 2012). Lensing has been successfully applied to constrain cosmological parameters via, for example, time delays in strong lenses (e.g. Refsdal 1964; Courbin et al. 2011; Suyu et al. 2010, 2013), multiple lensed source populations in galaxy clusters (e.g. Jullo et al. 2010), cosmic

shear due to large-scale structure (e.g. Schrabback et al. 2010; Kilbinger et al. 2013), and abundance and mass of galaxy clusters (e.g. Benson et al. 2013; von der Linden et al. 2014; Applegate et al. 2014). Highly magnifying gravitational lenses, such as a galaxy cluster, can also be used as nature's cosmic telescope to study the highest-redshift galaxies (e.g. Bradač et al. 2009; Coe et al. 2013). Studies of quasar microlensing have provided information about the quasar accretion disc size (e.g. Morgan et al. 2008; Mosquera et al. 2013), and Galactic microlensing has yielded numerous detections of planets (e.g. Beaulieu et al. 2006; Bachelet et al. 2012). Gravitational lensing has become a powerful method to study our Universe on a wide range of scales from planets to galaxies to galaxy clusters and to large-scale structures. We refer the reader to later chapters in this volume and to review articles (e.g. Treu et al. 2010; Bartelmann 2010; Limousin et al. 2013) for more details on the recent scientific results from lensing.

1.3. Basic lensing theory

In this section, we go through basic lensing theory in detail. We present the lens equation in Section 1.3.1, the Fermat potential and its connection to the time delays in Section 1.3.2, magnification and shear of lensed images in Section 1.3.3, classification of images in Section 1.3.4, odd-number and magnification theorems in Section 1.3.5, critical curves and caustics in Section 1.3.6, and the mass-sheet degeneracy in Section 1.3.7. Throughout this chapter, boldface symbols denote vectors, and regular symbols denote scalars.

1.3.1 *The lens equation*

We show in Figure 1.8 a typical lensing diagram. Physical coordinates (in units of, for example, Mpc) on the source plane are denoted by $\boldsymbol{\eta} = (\eta_1, \eta_2)$ and on the image plane are denoted by $\boldsymbol{\xi} = (\xi_1, \xi_2)$. The angular diameter distance between the observer and the lens (a.k.a. deflector) is D_{d}, between the observer and the source is D_{s} and between the lens and the source is D_{ds}. The light ray emanating from the source is deflected by an angle $\hat{\boldsymbol{\alpha}}$. Based on simple geometry and the definition of the angular diameter distance, the source and lens coordinates are related to the deflection angle $\hat{\boldsymbol{\alpha}}$ via

$$\boldsymbol{\eta} = \frac{D_{\mathrm{s}}}{D_{\mathrm{d}}}\boldsymbol{\xi} - D_{\mathrm{ds}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{1.4}$$

The angular coordinates on the source and lens planes are, respectively, $\boldsymbol{\beta} = \boldsymbol{\eta}/D_{\mathrm{s}}$ and $\boldsymbol{\theta} = \boldsymbol{\xi}/D_{\mathrm{d}}$. Using these in the above equation leads to

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{1.5}$$

where $\boldsymbol{\alpha}$ is the *scaled deflection angle* which is related to the true deflection angle $\hat{\boldsymbol{\alpha}}$ by

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\hat{\boldsymbol{\alpha}}(D_{\mathrm{d}}\boldsymbol{\theta}). \tag{1.6}$$

Equation (1.5) is known as the lens equation and it governs the deflection of light rays.

Recall from Section 1.2.1 that the general relativistic prediction of the deflection of a light ray due to a point mass M is

$$\hat{\boldsymbol{\alpha}} = \frac{4GM}{c^2\boldsymbol{\xi}}. \tag{1.7}$$

For a weak gravitational field and small deflection angles (i.e. a geometrically thin lens), a light ray described by the affine parameter λ with spatial trajectory $\boldsymbol{r} = (\xi_1(\lambda), \xi_2(\lambda),$
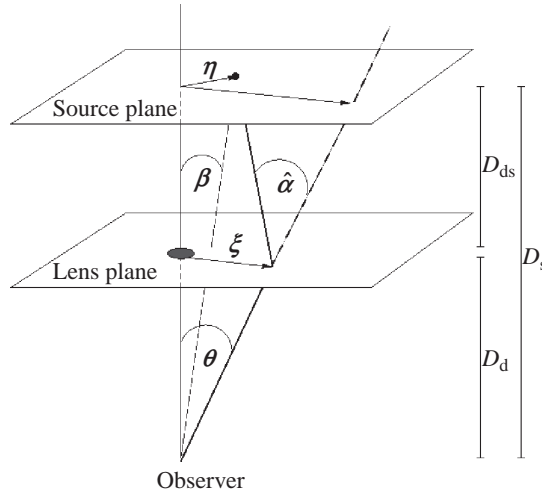
FIGURE 1.8. Schematic of lensing. Light from the source plane at coordinate $\boldsymbol{\eta}$ propagates to the lens plane at coordinate $\boldsymbol{\xi}$, gets deflected by angle $\hat{\boldsymbol{\alpha}}$ and arrives at the observer. The angular diameter distances are denoted by D_{d}, D_{s} and D_{ds} for the distance between the observer and the lens, the observer and the source, and the lens and the source, respectively. The observed angular position of the images, $\boldsymbol{\theta}$, is related to the angular position of the source, $\boldsymbol{\beta}$, via the lens equation (1.5). (Reproduced, with the permission of P. Schneider.)

$r_3(\lambda)$) that passes through a mass distribution with three-dimensional density $\rho(\boldsymbol{r'})$ will be deflected by

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \int \mathrm{d}r'_3\, \rho(\xi'_1, \xi'_2, r'_3) \frac{\boldsymbol{\xi} - \boldsymbol{\xi'}}{|\boldsymbol{\xi} - \boldsymbol{\xi'}|^2}. \tag{1.8}$$

Defining the *surface mass density* which corresponds to the projected mass density on to the lens plane,

$$\Sigma(\boldsymbol{\xi}) = \int \mathrm{d}r_3\, \rho(\xi_1, \xi_2, r_3), \tag{1.9}$$

we can rewrite equation (1.8) as

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi'\, \Sigma(\boldsymbol{\xi'}) \frac{\boldsymbol{\xi} - \boldsymbol{\xi'}}{|\boldsymbol{\xi} - \boldsymbol{\xi'}|^2}. \tag{1.10}$$

Combining equations (1.6) and (1.10), we obtain the following expression for the scaled deflection angle:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathcal{R}e^2} \mathrm{d}^2\theta'\, \kappa(\boldsymbol{\theta'}) \frac{\boldsymbol{\theta} - \boldsymbol{\theta'}}{|\boldsymbol{\theta} - \boldsymbol{\theta'}|^2}, \tag{1.11}$$

where $\kappa(\boldsymbol{\theta})$ is the *dimensionless surface mass density* (a.k.a. *convergence*),

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(D_{\mathrm{d}}\boldsymbol{\theta})}{\Sigma_{\mathrm{cr}}}, \tag{1.12}$$

and Σ_{cr} is the *critical surface mass density*,

$$\Sigma_{\mathrm{cr}} = \frac{c^2}{4\pi G} \frac{D_{\mathrm{s}}}{D_{\mathrm{d}}\, D_{\mathrm{ds}}}. \tag{1.13}$$

The critical density crudely distinguishes strong lenses from weak lenses. As we will see in Section 1.3.5, a lens with $\kappa > 1$ at some position $\boldsymbol{\theta}$ (corresponding to $\Sigma > \Sigma_{\mathrm{cr}}$) will be able to produce multiple images of a background source.

By using the property $\nabla \ln |\boldsymbol{\theta}| = \boldsymbol{\theta}/|\boldsymbol{\theta}^2|$, we can rewrite the scaled deflection angle $\boldsymbol{\alpha}(\boldsymbol{\theta})$ in terms of a scalar potential $\psi(\boldsymbol{\theta})$:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla\psi(\boldsymbol{\theta}), \tag{1.14}$$

where

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathcal{R}e^2} \mathrm{d}^2\theta' \kappa(\boldsymbol{\theta}') \ln |\boldsymbol{\theta} - \boldsymbol{\theta}'|. \tag{1.15}$$

This *lens potential* is related to the dimensionless surface mass density by the Poisson equation:

$$\nabla^2\psi = 2\kappa. \tag{1.16}$$

Therefore, given the lens potential ψ, one can easily compute the scaled deflection angle and the dimensionless surface mass density through differentiations.

1.3.2 *The Fermat potential and time delays*

The *Fermat potential* is defined as

$$\tau(\boldsymbol{\theta};\boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta}), \tag{1.17}$$

which is a function of $\boldsymbol{\theta}$ with $\boldsymbol{\beta}$ as a parameter. The Fermat potential is closely connected to the excess time delay, t, relative to the case of no lensing:

$$t(\boldsymbol{\theta};\boldsymbol{\beta}) = \frac{D_{\mathrm{d}}D_{\mathrm{s}}}{c\,D_{\mathrm{ds}}}(1 + z_{\mathrm{d}})\tau(\boldsymbol{\theta};\boldsymbol{\beta}) \tag{1.18}$$

$$= \frac{D_{\mathrm{d}}D_{\mathrm{s}}}{c\,D_{\mathrm{ds}}}(1 + z_{\mathrm{d}})\left[\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta})\right] \tag{1.19}$$

(for the derivation see, for example Schneider 1985; Schneider, Ehlers and Falco 1992). The first term in the square brackets corresponds to the geometric delay due to the additional path length of the observed light ray, and the second term is the gravitational delay associated with the lens mass distribution.

Fermat's principle states that rays of light traverse the path of stationary optical length with respect to variations of the path. In other words, we see lensed images of the source at positions $\boldsymbol{\theta}$, where $\nabla_{\boldsymbol{\theta}} t(\boldsymbol{\theta};\boldsymbol{\beta}) = \nabla_{\boldsymbol{\theta}}\tau(\boldsymbol{\theta};\boldsymbol{\beta}) = 0$. Note that applying this condition yields precisely the lens equation (1.5).

In Figure 1.9, taken from Blandford and Narayan (1986), we see the contours of the Fermat potential (or, equivalently, the time-delay surface given the constant proportionality between τ and t) for a simple elliptical lens with increasing surface density from panels a to d. The source is located at the centre of each diagram. In panel a, the lens has zero mass, and we see that the Fermat potential is symmetric about the centre with a clear minimum (marked 'L', signalling a low/minimum). In this case, we would see the single image of the source at the centre without lensing, as expected. As the lens surface mass density increases, the Fermat potential surface starts to get distorted, as seen in panel b. When the lens reaches sufficient mass density in panel c, an additional extremum and a saddle occur in the Fermat potential surface, corresponding to additional lensed images of the background source. The three images are of different types in this case – one is a minimum ('L' for low), one is a maximum ('H' for high), and the last one is a
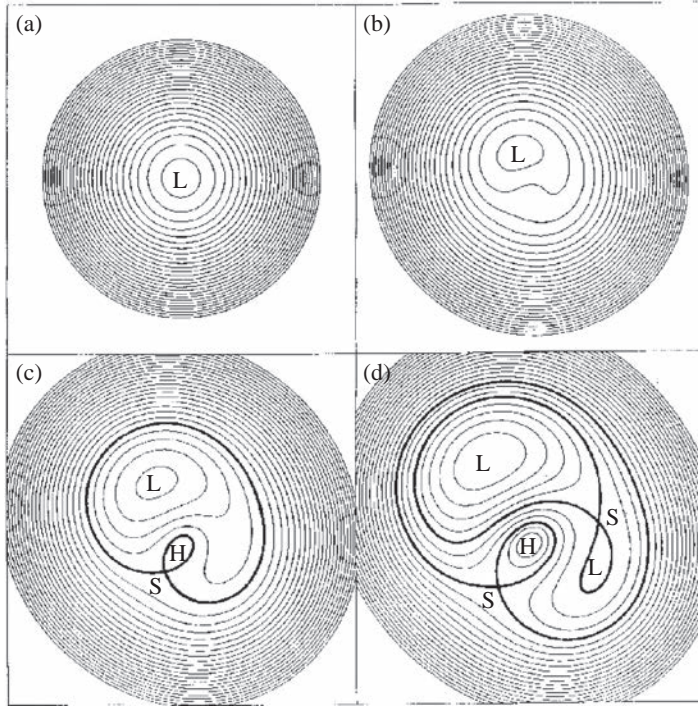
FIGURE 1.9. Fermat potential/time-delay surface for an elliptical lens galaxy with increasing surface mass density from panels a to d. As the lens surface mass density increases from zero (panel a) to higher and higher values (panels b–d), the Fermat potential surface gets distorted and additional extrema/saddles start to appear corresponding to additional multiple images of the background source. 'L' denotes a low (minimum), 'S' denotes a saddle, and 'H' denotes a high (maximum) in the surface. (Reproduced with permission from Blandford & Narayan 1986, *ApJ*, **310**, 568.)

saddle ('S'). When the lens mass density is increased even further, one can get another pair of saddle and minimum created in the Fermat potential. This lens system would thus have five images: two correspond to the minima in the Fermat potential, two to the saddles of the Fermat potential, and one to the maximum in the Fermat potential. We note that the central image corresponding to the maximum is often demagnified, especially for lens systems with cuspy profiles. Therefore, strong lens systems usually have two or four lensed images detected in observations, and are frequently referred to as 'doubles' and 'quads', respectively.

We can simplify equation (1.18) as

$$t(\boldsymbol{\theta}; \boldsymbol{\beta}) = \frac{D_{\Delta t}}{c} \tau(\boldsymbol{\theta}; \boldsymbol{\beta}), \qquad (1.20)$$

where $D_{\Delta t}$ is the so-called *time-delay distance*,

$$D_{\Delta t} \equiv (1 + z_{\mathrm{d}}) \frac{D_{\mathrm{d}} D_{\mathrm{s}}}{D_{\mathrm{ds}}}. \qquad (1.21)$$

For lens systems whose source brightnesses vary in time (e.g. galaxies that contain active galactic nuclei, AGN), one can monitor the brightnesses of the lensed images over

time and hence measure the time delay, Δt_{ij}, between the images at positions $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$:

$$\Delta t_{ij} \equiv t(\boldsymbol{\theta}_i, \boldsymbol{\beta}) - t(\boldsymbol{\theta}_j, \boldsymbol{\beta})$$
$$= \frac{D_{\Delta t}}{c} \left[\frac{(\boldsymbol{\theta}_i - \boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}_i) - \frac{(\boldsymbol{\theta}_j - \boldsymbol{\beta})^2}{2} + \psi(\boldsymbol{\theta}_j) \right]. \tag{1.22}$$

By using the image configuration and morphology, one can model the mass distribution of the lens to determine the lens potential $\psi(\boldsymbol{\theta})$ and the unlensed source position $\boldsymbol{\beta}$. Lens systems with measured time delays can therefore be used to determine $D_{\Delta t}$ via equation (1.22) and constrain cosmological models via the distance–redshift test (e.g. Refsdal 1964, 1966; Fadely et al. 2010; Courbin et al. 2011; Suyu et al. 2010, 2013). Having dimensions of distance, $D_{\Delta t}$ is inversely proportional to the Hubble constant H_0 and, being a combination of three angular diameter distances, it depends weakly on the other cosmological parameters as well.

1.3.3 *Image distortions: magnification and shear*

Let us consider a source that has an intrinsic flux $F_{\text{intrinsic}}$ (in the absence of lensing), and that is observed to have flux F_{observed} in the presence of a foreground lens. The magnification of the source is $F_{\text{observed}}/F_{\text{intrinsic}}$. By definition, the flux of the source is $F = I_\nu \times d\Omega$, where I_ν is the surface brightness and $d\Omega$ is the solid angle that the source subtends. Gravitational lensing conserves surface brightness as a result of Liouville's theorem, so the surface brightness of the source is the same with and without being lensed.[†] Therefore, the magnification of the source is simply

$$\mu = \frac{F_{\text{observed}}}{F_{\text{intrinsic}}} = \frac{d\Omega_{\text{observed}}}{d\Omega_{\text{intrinsic}}}. \tag{1.23}$$

To quantify the change in the size of the source, we define the Jacobian matrix

$$\mathcal{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}}, \tag{1.24}$$

which has components

$$\mathcal{A}_{ij} = \frac{\partial \beta_i}{\partial \theta_j}. \tag{1.25}$$

The *magnification factor* is then

$$\mu(\boldsymbol{\theta}) = \frac{1}{\det \mathcal{A}(\boldsymbol{\theta})}. \tag{1.26}$$

Images with $\mu > 0$ are referred to as positive parity images; likewise, images with $\mu < 0$ have negative parity and are mirror images of the background source. Locations where $\det \mathcal{A} = 0$ correspond to critical points/curves. The magnification diverges at these locations. This is because the geometric optics approximation breaks down in this regime, and wave optics must be used instead. The magnification of the source computed via wave optics will be finite, though potentially very high.

Applying the lens equation (1.5) to equation (1.24), we obtain

$$\mathcal{A}(\boldsymbol{\theta}) = \begin{pmatrix} 1 - \kappa(\boldsymbol{\theta}) - \gamma_1(\boldsymbol{\theta}) & -\gamma_2(\boldsymbol{\theta}) \\ -\gamma_2(\boldsymbol{\theta}) & 1 - \kappa(\boldsymbol{\theta}) + \gamma_1(\boldsymbol{\theta}) \end{pmatrix}, \tag{1.27}$$

[†] Provided there are no additional processes, such as dust extinction, that create or absorb the photons.

where γ_1 and γ_2 are the two components of *shear*,

$$\gamma \equiv \gamma_1 + \mathrm{i}\gamma_2 = |\gamma|\mathrm{e}^{2\mathrm{i}\varphi}. \tag{1.28}$$

Similar to the convergence, the shear components are second derivatives of the lens potential:

$$\gamma_1 = \frac{1}{2}(\psi_{,11} - \psi_{,22}), \tag{1.29}$$

$$\gamma_2 = \psi_{,12}, \tag{1.30}$$

where the subscripts ',1' and ',2' correspond to derivatives with respect to θ_1 and θ_2, respectively. The magnification in terms of the convergence and shear is thus

$$\mu = \frac{1}{\det \mathcal{A}} = \frac{1}{(1-\kappa)^2 - |\gamma|^2}. \tag{1.31}$$

The conservation of surface brightness can be expressed as

$$I(\boldsymbol{\theta}) = I^{(\mathrm{s})}[\boldsymbol{\beta}(\boldsymbol{\theta})], \tag{1.32}$$

where I and $I^{(\mathrm{s})}$ are the surface brightness of the image and the source, respectively. Based on this, we can visualize the distortion of lensed images by considering the locally linearized lens equation around the source position $\boldsymbol{\beta}_0$ with the corresponding image position $\boldsymbol{\theta}_0$:

$$\boldsymbol{\beta} - \boldsymbol{\beta}_0 = \mathcal{A}(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{1.33}$$

In particular, let us consider a small circular source with radius R that is parametrized by

$$\boldsymbol{\beta}(\lambda) = \boldsymbol{\beta}_0 + R(\cos\lambda, \sin\lambda), \tag{1.34}$$

where λ varies from 0 to 2π. By using the previous two equations, one can derive that the corresponding image of the small (infinitesimal) source is an ellipse with a major axis of

$$\frac{R}{1-\kappa-|\gamma|} = \frac{R}{(1-\kappa)(1-|g|)} \tag{1.35}$$

and a minor axis of

$$\frac{R}{1-\kappa+|\gamma|} = \frac{R}{(1-\kappa)(1+|g|)}, \tag{1.36}$$

where g is the *reduced shear* defined by

$$g(\boldsymbol{\theta}) = \frac{\gamma(\boldsymbol{\theta})}{1-\kappa(\boldsymbol{\theta})}. \tag{1.37}$$

The angle of the major axis of the image ellipse, as measured counterclockwise from θ_1, is the same as the shear angle φ in equation (1.28). Figure 1.10 illustrates this image distortion. The axis ratio of the ellipse is then

$$\frac{b}{a} = \frac{R}{(1-\kappa)(1+|g|)} \Big/ \frac{R}{(1-\kappa)(1-|g|)} \tag{1.38}$$

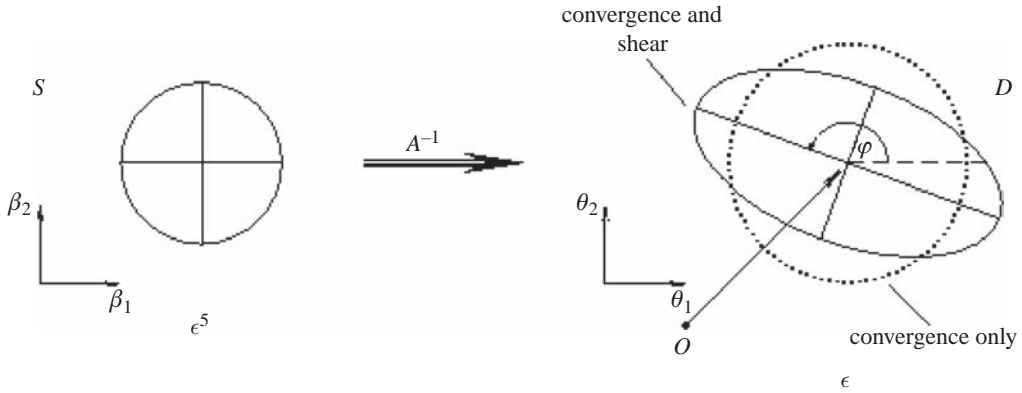$$= \frac{1-|g|}{1+|g|}. \tag{1.39}$$

FIGURE 1.10. The distortion of an infinitesimal background source when it is gravitationally lensed. A small circular source will be lensed into an ellipse with axis ratio $(1 - |g|)/(1 + |g|)$, where g is the reduced shear. (Reproduced, with the permission of M. Bradač, from Schneider et al. (eds) 2006, *Gravitational Lensing: Strong, Weak and Micro* (Berlin: Springer-Verlag).)

Rearranging this equation, we obtain an expression for the reduced shear in terms of the measured shape of the lensed source,

$$|g| = \frac{1 - b/a}{1 + b/a}. \tag{1.40}$$

Therefore, if a background source is intrinsically round, then, by measuring its lensed shape (i.e. the axis ratio of the elliptical image), we can infer the reduced shear. However, source galaxies are typically not intrinsically round and the distortion due to the shear is typically smaller than the intrinsic ellipticities of the sources. To overcome this, one could assume that the intrinsic ellipticities are *randomly orientated* and average the measured shapes over many different sources to measure the reduced shear in the weak lensing regime.

Once we have shear measurements, we can then infer the mass distribution since both the convergence and shear are derived from the same lens potential. Using equations (1.29), (1.30) and (1.15), we can write the shear as

$$\gamma(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathcal{R}e^2} d^2\theta' \, \mathcal{D}(\boldsymbol{\theta} - \boldsymbol{\theta}') \, \kappa(\boldsymbol{\theta}'), \tag{1.41}$$

where

$$\mathcal{D}(\boldsymbol{\theta}) \equiv -\frac{\theta_1^2 - \theta_2^2 + 2i\theta_1\theta_2}{|\boldsymbol{\theta}|^4} = \frac{-1}{(\theta_1 - i\theta_2)^2}. \tag{1.42}$$

By inverting this integral relation, one finds that

$$\kappa(\boldsymbol{\theta}) - \kappa_0 = \frac{1}{\pi} \int_{\mathcal{R}e^2} d^2\theta' \, \mathcal{R}e \left[\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}') \, \gamma(\boldsymbol{\theta}') \right], \tag{1.43}$$

where κ_0 is a constant. Its physical origin is that a constant sheet of mass produces no shear. This 'mass-sheet degeneracy' will be discussed in more detail in Section 1.3.7. This integral equation forms the basis of mass reconstruction from weak lensing. By measuring the shear of background sources, one can solve the above integral equation to derive the convergence of the mass distributions.

1.3.4 *Classification of images*

Recall from Section 1.3.2 that ordinary images (where $\det \mathcal{A} \neq 0$) occur at $\nabla_{\boldsymbol{\theta}} \tau(\boldsymbol{\theta}; \boldsymbol{\beta}) = 0$. Note also that $\tau_{ij} = \mathcal{A}_{ij}$ by construction. We define three types of ordinary images:

- Type I: minimum of τ
    $\det \mathcal{A} > 0$ and $\operatorname{tr} \mathcal{A} > 0$
- Type II: saddle point of τ
    $\det \mathcal{A} < 0$
- Type III: maximum of τ
    $\det \mathcal{A} > 0$ and $\operatorname{tr} \mathcal{A} < 0$

The bottom right-hand panel of Figure 1.9 shows the different types of images, with 'L' corresponding to Type I, 'S' corresponding to Type II, and 'H' corresponding to Type III. Type II and III images occur when lensing becomes strong. Even if the lens is not a strong one, there is at least a Type I image for lens mass distributions that are smooth and of finite total mass.

1.3.5 *Theorems*

We state here two theorems concerning the number and magnification of images, and the implications of these theorems.

Odd-number theorem. Given a smooth lens mass distribution $\kappa(\boldsymbol{\theta})$ that decreases faster than $|\boldsymbol{\theta}|^{-2}$ for $|\boldsymbol{\theta}| \to \infty$, the lens has finite total mass, and the deflection angle field $\boldsymbol{\alpha}(\boldsymbol{\theta})$ is continuous and bounded. Let the number of Type I, II, and III images be denoted by n_{I}, n_{II}, and n_{III}, respectively, and the total number of images be $n \equiv n_{\mathrm{I}} + n_{\mathrm{II}} + n_{\mathrm{III}}$. For a source at position $\boldsymbol{\beta}$ that does not correspond to where $\det \mathcal{A} = 0$, the following properties hold:

(i) $n_{\mathrm{I}} \geq 1$
(ii) $n < \infty$
(iii) $n_{\mathrm{I}} + n_{\mathrm{III}} = 1 + n_{\mathrm{II}}$
(iv) For sufficiently large $\boldsymbol{\beta}$, $n = n_{\mathrm{I}} = 1$

These properties imply that

- the total number of images $n = 1 + 2n_{\mathrm{II}}$ is odd (the so-called 'odd-number theorem')
- the number of positive-parity images (Type I and III) exceeds that of negative-parity images (Type II) by 1
- $n > 1$ if and only if $n_{\mathrm{II}} > 1$

For the proof of these, we refer the reader to Burke (1981) and Schneider et al. (1992). The odd-number theorem and its corollaries can be understood by looking at Figure 1.9. There will always be a Type I image of the source, the limiting case in the absence of a lens. In order to create more images, a saddle (Type II image) in the time-delay surface needs to form with an additional minimum (Type I) or maximum (Type III) image. Hence, images are created in pairs, so the total number must be odd. This also explains why, if there are multiple images, at least one of them must be a saddle (Type II).

Magnification theorem. Under the same conditions as for the odd-number theorem (i.e. $\kappa(\boldsymbol{\theta})$ is smooth and decreases faster than $|\boldsymbol{\theta}|^{-2}$ for $|\boldsymbol{\theta}| \to \infty$, and a source at position $\boldsymbol{\beta}$ which does not correspond to where $\det \mathcal{A} = 0$), then that image of the source which arrives first at the observer is of Type I and appears brighter than, or equally bright as, the source would appear in the absence of the lens (Schneider 1984).
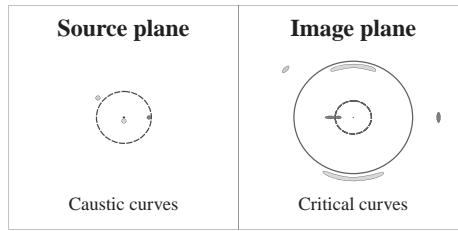
FIGURE 1.11. Critical curves (right) and caustics (left) of a non-singular isothermal sphere lens mass distribution. The solid critical curve maps into the central point on the source plane, whereas the dashed critical curve maps into the dashed caustics. The source plane (left) shows three sources in different shades and positions, and the respective images of these sources are shown on the image plane (right) in corresponding shades. For a source situated outside the dashed caustics, there is only one image of the source. Inside the dashed caustics, there are three images, with a highly demagnified central image for the light set of images. (Figure courtesy of Adam Amara.)

Two corollaries of these theorems are:
1. An isolated transparent lens can produce multiple images if and only if there is a point $\boldsymbol{\theta}$ with $\det \mathcal{A} < 0$.
2. A sufficient (but not necessary) condition for possible multiple images is that there exists a point $\boldsymbol{\theta}$ such that $\kappa(\boldsymbol{\theta}) > 1$.

1.3.6 *Critical and caustic curves*

As already mentioned briefly in Section 1.3.3, the locations on the image plane $\boldsymbol{\theta}$ where $\det \mathcal{A} = 0$ are called *critical curves*. These correspond to locations of high magnification. By mapping the critical curves to the source plane using the lens equation (1.5), we obtain the *caustics* on the source plane. In Figure 1.11, we show the caustics (left-hand panel) and the critical curves (right-hand panel) for a non-singular isothermal sphere[†] lens mass density described by $\rho(r) = \rho_0 \, r_{\mathrm{c}}^2 / (r_{\mathrm{c}}^2 + r^2)$, where r is the three-dimensional radius, ρ_0 is a normalization, and r_{c} is the core radius. The dashed critical curve maps into the dashed caustics, whereas the solid critical curve maps into the central point on the source plane. As a result of the axisymmetry in this system, if the source is perfectly aligned with the observer and lens (i.e. the source is located at the central caustic point), the source will be lensed into a full ring. The solid critical curve is known as the *tangential* critical curve since the multiple lensed images straddle tangentially around this curve. The dashed critical curve is known as the *radial* critical curve since the multiple images are magnified radially in the vicinity of the radial critical curve.

When the lens mass distribution deviates from spherical symmetry, the caustic structure changes and additional images can appear. Let us consider an elliptical mass distribution where the projected mass density κ is a function of $\theta_1^2 + \theta_2^2/q^2$, (θ_1, θ_2) are the coordinates on the image plane, and q is the axis ratio of the isodensity contours. When $q = 1$, we have the spherical isothermal lens mass distribution that was discussed earlier. Shown in Figure 1.12 are the caustics and critical curves for a non-singular isothermal ellipsoid lens. The caustic as a central point in Figure 1.11 for the spherical case becomes a diamond shaped caustic, called an *astroid*, when κ is elliptical. The astroid has four smooth curves known as *folds* that join at the four *cusps*.

[†] The so-called isothermal profile leads to a flat rotation curve.

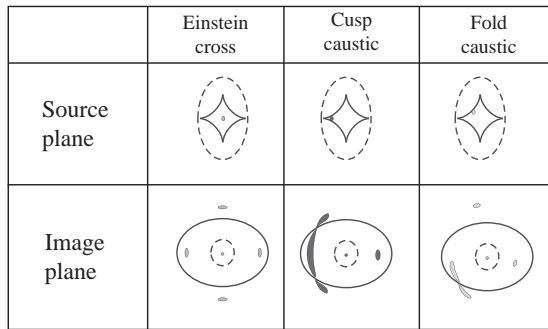|	Einstein cross	Cusp caustic	Fold caustic
Source plane			
Image plane			

FIGURE 1.12. Caustics (top row) and critical curves (bottom row) of a non-singular isothermal ellipsoid lens mass distribution. The large solid tangential critical curve maps into the astroid caustic, and the small dashed radial critical curve maps into the large caustics on the source plane. Within the astroid, five multiple images of the source are formed; they can be one of three generic image configurations, 'Einstein cross', 'cusp', and 'fold', depending on the position of the source relative to the astroid caustic. (Figure courtesy of Adam Amara.)

The caustics separate regions of different image multiplicity. A source lying outside the outer dashed caustic will have one image, a source lying inside the dashed caustic but outside the solid astroid will be lensed into three multiple images, and a source within the astroid will produce five images. The columns in the figure show the three generic configurations of image systems that one can get with different source positions within the astroid. When the source is nearly aligned with the observer and the lens, an 'Einstein cross' configuration is formed, where four of the five images are distributed nearly symmetrically around the lens centre. When the source lies near a cusp, three of the multiple images will be close to one another on the image plane forming the 'cusp' configuration. When the source lies near a fold, two of the multiple images will be close to each other on the image plane forming the 'fold' configuration. In all cases, there is a fifth image near the centre of the lens.

Most of the observed galaxy-scale gravitational lens systems to date have two or four lensed images, as opposed to the three and five described above. How is this compatible with the odd-number theorem? The reason for the one fewer image observed is that the image near the centre is often demagnified as a result of the high central surface mass density of the lens. The smaller the core radius of the lens mass profile, the more demagnified is the central image. In fact, radio observations of central images in lenses have been used to place constraints on the mass of the black hole at the centre of the lens galaxy, as well as the galaxy's central surface mass density (e.g. Winn, Rusin and Kochanek 2004). Central images are much more difficult to detect in optical observations of lenses, since lens galaxies are often optically bright and outshine the highly demagnified central images.

For smooth lens mass distributions, the merging images in the fold and cusp lenses obey relations concerning their fluxes. In particular, in a fold lens with the two merging images on opposite sides of the critical curve labelled A and B, the magnifications of images A and B should be of the same magnitude but opposite parity. This is known as the fold relation:

$$\mu_A + \mu_B = 0, \tag{1.44}$$

$$|\mu_A| = |\mu_B|. \tag{1.45}$$

In the case of cusp configurations, with image B labelled as the image on one side of the tangential critical curve, and images A and C as nearby images on the other side
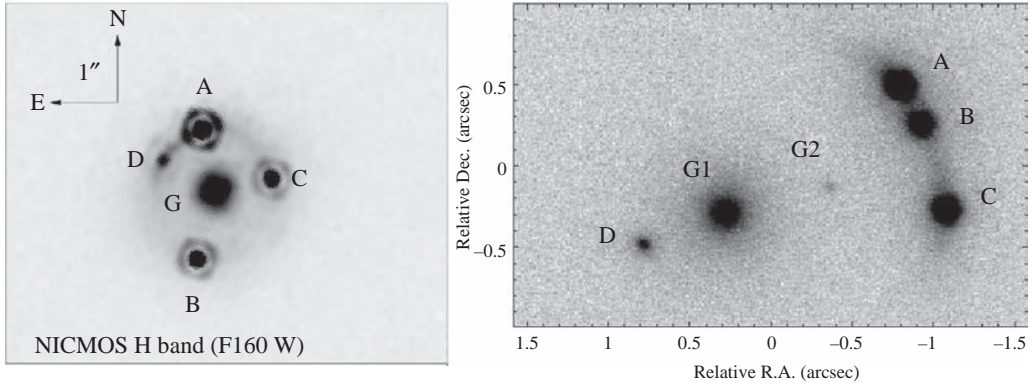
FIGURE 1.13. Examples of lenses that violate the fold or cusp relations. Left: an *HST* image of the lens system 0924+0219 showing the two merging images, A and D, in the fold lens having drastically different fluxes. (Figure taken from Morgan et al. 2006 *ApJ*, **647**, 874.) Right: an adaptive-optics image from Keck Observatory of the lens system 2045+265 showing that the three merging images have image B fainter than A and C instead of being the sum of the fluxes of A and C. (Reproduced with permission from McKean et al. 2007, *MNRAS*, **378**, 109.)

of the critical curve (see Figure 1.12 bottom-middle panel), the relations among the magnifications of the three images are

$$\mu_{\rm A} + \mu_{\rm B} + \mu_{\rm C} = 0, \tag{1.46}$$

$$|\mu_{\rm B}| = |\mu_{\rm A}| + |\mu_{\rm C}|. \tag{1.47}$$

However, in reality, these relations are often violated, as shown by the examples of fold and cusp lenses in Figure 1.13. Possible causes of these violations are (1) variability in the source surface brightness and the time delay between the images, (2) differential dust extinction between the images, (3) microlensing of the images by stars in the lens galaxy, and (4) the presence of dark matter substructure in the lens galaxy. By eliminating the first three causes, lensing provides a unique way to detect and probe dark matter substructures (e.g. Dalal & Kochanek 2002; Vegetti et al. 2012).

1.3.7 *Mass-sheet degeneracy*

As we have seen, gravitational lensing provides a way to measure the mass distribution of the lens galaxies. Nonetheless, given a model of the lens mass distribution, there is a family of solutions that can describe typical lensing observations equally well. This is the so-called 'mass-sheet degeneracy' in gravitational lensing, which was first noticed by Falco, Gorenstein and Shapiro (1985) and Gorenstein, Shapiro and Falco (1988). In this section, we explain the mass-sheet degeneracy and describe ways to break this degeneracy.

Suppose we have a lens mass distribution $\kappa(\boldsymbol{\theta})$ with corresponding lens potential $\psi(\boldsymbol{\theta})$. Let us consider the following transformation:

$$\psi_\lambda(\boldsymbol{\theta}) = \frac{\lambda}{2}|\boldsymbol{\theta}|^2 + \boldsymbol{s} \cdot \boldsymbol{\theta} + c_0 + (1 - \lambda)\psi(\boldsymbol{\theta}), \tag{1.48}$$

where λ and c_0 are constant scalars and $\boldsymbol{s}$ is a constant vector. Applying equation (1.14), we obtain the transformed deflection angle as

$$\boldsymbol{\alpha}_\lambda(\boldsymbol{\theta}) = \lambda\boldsymbol{\theta} + \boldsymbol{s} + (1 - \lambda)\boldsymbol{\alpha}(\boldsymbol{\theta}). \tag{1.49}$$

The constant c_0 in $\psi_\lambda(\boldsymbol{\theta})$ drops out in the differentiation, so this zero-point in the lens potential is unobservable and has no physical significance. The $\boldsymbol{s} \cdot \boldsymbol{\theta}$ term in $\psi_\lambda(\boldsymbol{\theta})$
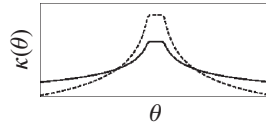
FIGURE 1.14. Illustration of the mass-sheet transformation. The dashed line is the original lens mass distribution $\kappa(\boldsymbol{\theta})$ and the solid line is the transformed distribution $\kappa_\lambda(\boldsymbol{\theta})$ in equation (1.50).

becomes $\boldsymbol{s}$ in $\boldsymbol{\alpha}_\lambda(\boldsymbol{\theta})$, which corresponds to a constant shift on the source plane that also has no physical significance since we do not observe the unlensed source directly. Therefore, the crucial terms in equation (1.48) are the first and last term on the right-hand side of the equation. Using equation (1.16), the transformed convergence is

$$\kappa_\lambda(\boldsymbol{\theta}) = \lambda + (1 - \lambda)\kappa(\boldsymbol{\theta}). \tag{1.50}$$

This transformation is illustrated in Figure 1.14, where the dashed profile indicates the original $\kappa(\boldsymbol{\theta})$ and the solid profile is the transformed $\kappa_\lambda(\boldsymbol{\theta})$. The transformation renormalizes the original distribution by $(1 - \lambda)$ and adds a constant sheet of mass density λ, hence the name 'mass-sheet' degeneracy.

The lens equation for the transformed distribution is

$$\boldsymbol{\beta}_\lambda = \boldsymbol{\theta} - \boldsymbol{\alpha}_\lambda(\boldsymbol{\theta}) = \boldsymbol{\theta} - \lambda\boldsymbol{\theta} - \boldsymbol{s} - (1 - \lambda)\boldsymbol{\alpha}(\boldsymbol{\theta}). \tag{1.51}$$

Rearranging the above and using the lens equation $\boldsymbol{\theta} - \boldsymbol{\alpha} = \boldsymbol{\beta}$, we obtain

$$\boldsymbol{\beta}_\lambda = (1 - \lambda)\boldsymbol{\beta} - \boldsymbol{s}. \tag{1.52}$$

Comparing the original source coordinates $\boldsymbol{\beta}$ and the transformed coordinates $\boldsymbol{\beta}_\lambda$, we see that the transformation in equation (1.50) simply rescales the source by $(1 - \lambda)$ and shifts it by $\boldsymbol{s}$. The family of $\kappa_\lambda(\boldsymbol{\theta})$ solutions will describe the observed positions and shapes equally well by rescaling/shifting the source that is unobservable. If we know the intrinsic size of the source via other means, then we can break the degeneracy.

How about the fluxes of the lensed images? Based on equation (1.52), the Jacobian matrix scales as

$$\mathcal{A}_\lambda = (1 - \lambda)\mathcal{A}. \tag{1.53}$$

This implies that

$$\mu_\lambda = \frac{\mu}{(1 - \lambda)^2}, \tag{1.54}$$

so the relative fluxes of images are invariant; in other words, $\mu_\lambda^{\text{image 1}}/\mu_\lambda^{\text{image 2}} = \mu^{\text{image 1}}/\mu^{\text{image 2}}$. The absolute flux is not invariant, but unless we know the intrinsic brightness of the source the mass-sheet transformation remains a degeneracy.

Using the expression of $\mathcal{A}$ in equation (1.27) and the transformation in equation (1.53), we see that

$$\gamma_\lambda(\boldsymbol{\theta}) = (1 - \lambda)\gamma(\boldsymbol{\theta}), \tag{1.55}$$
$$(1 - \kappa_\lambda) = (1 - \lambda)(1 - \kappa). \tag{1.56}$$

Therefore, the reduced shear g defined in equation (1.37) is invariant under the mass-sheet transformation in equation (1.50).

Even though the entire family of κ_λ can fit the lensing observables, in practice, we know that mass distributions of galaxies do not have a constant slab of mass. There could be

some constant projected surface mass density due to nearby galaxy groups/clusters or large scale structure. Typically, the *external convergence*, which is frequently described by a single parameter κ_{ext} to characterize the mass density external to the strong lens, has magnitudes on the order of $\kappa_{\text{ext}} \sim 0.05$.[†] Therefore, for most of the astrophysical applications of lensing, the mass-sheet degeneracy for such values of κ_{ext} is of small concern, as it leads to only a small percentage uncertainty in the mass of the lens deflector and in the size/luminosity of the source.

While many of the lensing observables are invariant under the mass-sheet transformation, the time delay between the multiple images is not. Applying equations (1.48) and (1.52) to equation (1.17), we get that

$$\tau_\lambda(\boldsymbol{\theta}; \boldsymbol{\beta}_\lambda) = (1 - \lambda)\tau(\boldsymbol{\theta}; \boldsymbol{\beta}) + \text{constant}, \tag{1.57}$$

where the constant term on the right-hand side depends on the source position but not on the image position. If we compute the Fermat potential difference between two images, we obtain

$$\Delta\tau_\lambda(\boldsymbol{\theta}; \boldsymbol{\beta}_\lambda) = (1 - \lambda)\Delta\tau(\boldsymbol{\theta}; \boldsymbol{\beta}). \tag{1.58}$$

According to equation (1.20),

$$\Delta t(\boldsymbol{\theta}; \boldsymbol{\beta}) = \frac{D_{\Delta t}}{c}\Delta\tau(\boldsymbol{\theta}; \boldsymbol{\beta}), \tag{1.59}$$

which means that the time delays between the multiple images also scale as $(1 - \lambda)$. If we wish to use time-delay lenses for cosmography and determine $D_{\Delta t}$, then given measured (fixed) Δt between the images, this means that $D_{\Delta t}$ is scaled by

$$D_{\Delta t, \lambda} = \frac{D_{\Delta t}}{1 - \lambda}. \tag{1.60}$$

This is problematic for time-delay lens cosmography, where $D_{\Delta t}$ needs to be measured at the few per cent level. In fact, the time-delay distance measurements are currently limited by our knowledge of the mass structures along the line of sight to the source (e.g. Fadely et al. 2010; Suyu et al. 2010, 2013). In order to break this mass-sheet degeneracy, one needs additional information.

Two practical ways to break the mass-sheet degeneracy are (1) to use the stellar kinematics of the lens galaxy (e.g. Treu & Koopmans 2002; Koopmans & Treu 2003; Barnabè et al. 2009; Auger et al. 2010; Suyu et al. 2010; Sonnenfeld et al. 2012) to make an independent estimate of the lens mass, and (2) to study the environment of the lens system (e.g. Keeton & Zaludoff 2004; Fassnacht et al. 2006; Momcheva et al. 2006; Suyu et al. 2010; Wong et al. 2011; Fassnacht, Koopmans and Wong 2011) in order to estimate κ_{ext} directly. In addition, if sources at different redshifts are strongly lensed by the same deflector, one can partly break the mass-sheet degeneracy since the different $D_{\text{ds}}/D_{\text{s}}$ ratios of the sources imply that the mass-sheet transformation cannot leave both sets of the observables (positions, shapes, and relative fluxes) of the lensed sources invariant.

1.4. Simple lens mass distributions

In this section, we present simple yet useful lens mass distributions. We begin by describing properties of axisymmetric lens mass distributions in Section 1.4.1. We then

---

[†] We distinguish between λ and κ_{ext}; λ is a simple parameter for the mathematical mass-sheet transformation (equation 1.50) and λ can take on any value, whereas κ_{ext} is a parameter used to characterize the physical amount of external convergence in a lens system.

work through the singular isothermal sphere (SIS) model in Section 1.4.2. We mention briefly in Section 1.4.3 the two commonly used models with more flexibility: the non-singular isothermal sphere (NIS) model with external shear and the non-singular isothermal ellipsoid (NIE) model. Finally, we briefly describe the power-law lens mass distributions and their applications to astrophysics in Section 1.4.4.

1.4.1 *Axisymmetric mass distributions*

An axisymmetric mass distribution is one where κ depends only on the radial coordinate $|\boldsymbol{\theta}|$, i.e. $\kappa(\boldsymbol{\theta}) = \kappa(|\boldsymbol{\theta}|) = \kappa(\theta)$. Using equation (1.11), we obtain the following scaled deflection angle for axisymmetric lens distributions:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{2\boldsymbol{\theta}}{|\boldsymbol{\theta}|^2} \int_0^{|\boldsymbol{\theta}|} \mathrm{d}\theta'\, \theta'\, \kappa(\theta'). \tag{1.61}$$

Note that $\boldsymbol{\alpha}(\boldsymbol{\theta})$ is collinear with $\boldsymbol{\theta}$. The lens equation (1.5) implies that the source position $\boldsymbol{\beta}$ is also collinear with $\boldsymbol{\theta}$. Since the three quantities are collinear, the lens equation reduces to a one-dimensional equation. Defining $\boldsymbol{\beta} = \beta\hat{\boldsymbol{e}}$, we obtain $\boldsymbol{\theta} = \theta\hat{\boldsymbol{e}}$ and $\boldsymbol{\alpha} = \alpha\hat{\boldsymbol{e}}$, where

$$\alpha(\theta) = \frac{2}{\theta} \int_0^{\theta} \mathrm{d}\theta'\, \theta'\, \kappa(\theta'). \tag{1.62}$$

The corresponding one-dimensional lens equation is

$$\beta = \theta - \alpha(\theta). \tag{1.63}$$

Note that $\alpha(-\theta) = -\alpha(\theta)$ for axisymmetric lens distributions, as evident from equation (1.62).

We define the *mean surface mass density* inside circular radius θ as

$$\bar{\kappa}(\theta) = \frac{m(\theta)}{\theta^2}, \tag{1.64}$$

with the *dimensionless mass* inside θ as

$$m(\theta) = 2 \int_0^{\theta} \mathrm{d}\theta'\, \theta'\, \kappa(\theta'). \tag{1.65}$$

With these two definitions, we can rewrite the scaled deflection angle as

$$\alpha(\theta) = \frac{m(\theta)}{\theta} = \bar{\kappa}(\theta)\theta. \tag{1.66}$$

Therefore, the (vector) lens equation (1.5) can be written as

$$\boldsymbol{\beta} = [1 - \bar{\kappa}(|\boldsymbol{\theta}|)]\,\boldsymbol{\theta}. \tag{1.67}$$

From this, one can use the definition of the Jacobian matrix in equation (1.24) to derive

$$\det \mathcal{A} = (1 - \bar{\kappa})(1 + \bar{\kappa} - 2\kappa). \tag{1.68}$$

Recall that $\det \mathcal{A} = 0$ defines the location of the critical curves. Therefore, we have two possibilities for the critical curves: (1) *tangential critical curves* with $1 - \bar{\kappa} = 0$, or (2) *radial critical curves* with $1 + \bar{\kappa} - 2\kappa = 0$. The names of the critical curves originate from the distortion of a circular source as observed on the image plane – near the tangential critical curve, the images are distorted tangentially, whereas images near

the radial critical curve are distorted radially across the critical curve. We already saw examples of this in Figure 1.11: the light-shaded images are tangentially distorted along the solid tangential critical curve, whereas the dark-shaded images are distorted radially along the dashed radial critical curves.

The tangential critical curve at radius $\theta_{\rm E}$, also known as the *Einstein radius*, satisfies (by definition)

$$\bar{\kappa}(\theta_{\rm E}) = 1. \tag{1.69}$$

The mass enclosed within $\theta_{\rm E}$ is

$$M(\leq \theta_{\rm E}) = \bar{\kappa}(\theta_{\rm E})\,\pi\theta_{\rm E}^2\,D_{\rm d}^2\,\Sigma_{\rm cr} \tag{1.70}$$

$$= \pi\theta_{\rm E}^2\,D_{\rm d}^2\,\Sigma_{\rm cr}. \tag{1.71}$$

The Einstein radius is quite insensitive to model assumptions and can be measured accurately, especially in four-image systems. Strong lensing yields accurate enclosed mass measurements with uncertainties that are typically less than $\sim 5\%$ (e.g. Bolton et al. 2008). Rewriting equation (1.71), we can express the Einstein radius in terms of the enclosed mass:

$$\theta_{\rm E} = \left(\frac{4GM}{c^2}\frac{D_{\rm ds}}{D_{\rm d}D_{\rm s}}\right)^{1/2} \tag{1.72}$$

$$\approx 0.9'' \left(\frac{M(\leq \theta_{\rm E})}{10^{12}M_\odot}\right)^{1/2}\left(\frac{D_{\rm ds}\,1{\rm Gpc}}{D_{\rm d}D_{\rm s}}\right)^{1/2}. \tag{1.73}$$

Thus, we see that the mass scale sets the Einstein radius (tangential critical curves), which is approximately the locations of the tangential arcs from the lens centre. Galaxy-scale lenses typically have $\theta_{\rm E} \sim 1''$.

1.4.2 *Singular isothermal sphere*

The three-dimensional mass density of a singular isothermal sphere (SIS) is

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2}, \tag{1.74}$$

where r is the three-dimensional radius, σ_v is the velocity dispersion and G is the gravitational constant. This distribution leads to a flat rotation curve with rotation velocity $v_{\rm c} = \sqrt{2}\sigma_v$. The surface mass density of the SIS is

$$\Sigma(\xi) = \int_{-\infty}^{\infty} {\rm d}r_3\,\rho\left(\sqrt{\xi^2 + r_3^2}\right) = \frac{\sigma_v^2}{2G}\xi^{-1}. \tag{1.75}$$

Scaling it by $\Sigma_{\rm cr}$, we get the convergence:

$$\kappa(\theta) = \frac{\theta_{\rm E}}{2|\theta|}, \tag{1.76}$$

where

$$\theta_{\rm E} = 4\pi\left(\frac{\sigma_v}{c}\right)^2\frac{D_{\rm ds}}{D_{\rm s}}. \tag{1.77}$$
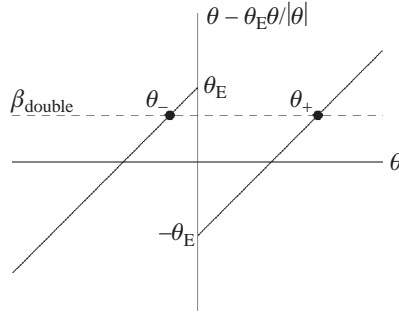
FIGURE 1.15. Image positions for the SIS lens. The thick solid lines show the source position β, given by the lens equation ($= \theta - \theta_{\mathrm{E}}\theta/|\theta|$), as a function of the image position θ. For source positions within $\pm\theta_{\mathrm{E}}$ (e.g. at β_{double} shown in the figure), there are two images of the source at θ_- and θ_+. The two images are separated by $2\theta_{\mathrm{E}}$.

One can derive the following properties of the SIS:

$$\bar{\kappa}(\theta) = \frac{\theta_{\mathrm{E}}}{|\theta|}, \tag{1.78}$$

$$|\gamma|(\theta) = \frac{\theta_{\mathrm{E}}}{2|\theta|}, \tag{1.79}$$

$$\text{and } \alpha(\theta) = \theta_{\mathrm{E}}\frac{\theta}{|\theta|}. \tag{1.80}$$

Note that $\bar{\kappa}(\theta_{\mathrm{E}}) = 1$ so θ_{E} in equation (1.77) corresponds to the tangential critical curve.

The one-dimensional lens equation for the SIS is

$$\beta = \theta - \theta_{\mathrm{E}}\frac{\theta}{|\theta|}. \tag{1.81}$$

To investigate the positions of the (possibly multiple) images, we plot β in equation (1.81) versus θ in Figure 1.15. As evident from this plot, for source positions $\beta \leq -\theta_{\mathrm{E}}$ or $\beta \geq \theta_{\mathrm{E}}$, there is only one image of the background source, at $\theta = \beta - \theta_{\mathrm{E}}$ or $\theta = \beta + \theta_{\mathrm{E}}$, respectively. For $-\theta_{\mathrm{E}} < \beta < \theta_{\mathrm{E}}$, there are two images of the source at $\theta_- = \beta - \theta_{\mathrm{E}}$ and $\theta_+ = \beta + \theta_{\mathrm{E}}$. The two images are therefore separated by $2\theta_{\mathrm{E}}$.

Using equation (1.26), one can derive the magnification at image position θ:

$$\mu = \frac{1}{\det \mathcal{A}} = \frac{|\theta|}{|\theta| - \theta_{\mathrm{E}}}. \tag{1.82}$$

Therefore, for the image θ_+, the corresponding μ_+ is always greater than 1. In other words, the image at θ_+ is always magnified. In contrast, the image at θ_- can be highly demagnified, particularly as $\theta_- \to 0$, or as $\beta \to \theta_{\mathrm{E}}$.

1.4.3 *Isothermal models beyond SIS*

The SIS model in the previous section can produce a maximum of two observable images of the background source (for an NIS model, a third central image exists, as shown in Figure 1.11). To create a four-image lens system that exists in reality, one has to deviate from isothermality or axisymmetry. In this section, we consider two frequently employed isothermal models that break the axisymmetry in κ: (1) NIS (non-singular isothermal sphere) in an external shear field and (2) NIE (non-singular isothermal ellipsoid).

1.4.3.1 NIS with an external shear

One form of expressing the NIS that reduces to SIS in the large θ limit is

$$\kappa(\theta) = \frac{\theta_{\mathrm{e}}}{\theta_{\mathrm{c}}} \left(1 + \frac{\theta^2}{2\theta_{\mathrm{c}}^2}\right) \left(1 + \frac{\theta^2}{\theta_{\mathrm{c}}^2}\right)^{-3/2}, \tag{1.83}$$

where θ_{e} characterizes the lens strength, and θ_{c} is the core radius. In contrast to the SIS, the above NIS has a finite $\kappa(0)$. The form of NIS above leads to a simple expression for $\bar{\kappa}$, namely,

$$\bar{\kappa}(\theta) = \frac{\theta_{\mathrm{e}}}{\sqrt{\theta^2 + \theta_{\mathrm{c}}^2}}, \tag{1.84}$$

so the deflection angle $\alpha = \bar{\kappa}\theta$ is also of a simple expression.

Using the criterion $\bar{\kappa} = 1$ (see equation 1.68), the tangential critical curve occurs at $\theta_{\mathrm{E}} = (\theta_{\mathrm{e}}^2 - \theta_{\mathrm{c}}^2)^{1/2}$ for $\theta_{\mathrm{c}} < \theta_{\mathrm{e}}$, and the lens is not critical for $\theta_{\mathrm{c}} > \theta_{\mathrm{e}}$. On the other hand, the radial critical curve occurs at $\theta_{\mathrm{r}} = \theta_{\mathrm{c}}\sqrt{(\theta_{\mathrm{e}}/\theta_{\mathrm{c}})^{2/3} - 1}$. Note that in the limit of $\theta_{\mathrm{c}} \to 0$, we recover the case of SIS: $\theta_{\mathrm{e}} \to \theta_{\mathrm{E}}$ and $\theta_{\mathrm{r}} \to 0$.

Mass structures along the line of sight to the source can induce an external shear on the lens system. One way to express the lens potential of a constant external shear in polar coordinates θ and φ (such that $(\theta_1, \theta_2) = (\theta\cos\varphi, \theta\sin\varphi)$ are the cartesian coordinates on the lens plane) is

$$\psi_{\mathrm{ext}}(\theta, \varphi) = \frac{1}{2}\gamma_{\mathrm{ext}}\theta^2 \cos 2(\varphi - \phi_{\mathrm{ext}}), \tag{1.85}$$

where γ_{ext} is the shear strength and ϕ_{ext} is the shear angle. Note that $\kappa_{\mathrm{ext}} = \frac{1}{2}\nabla^2\psi_{\mathrm{ext}}$ is zero. The shear position angle of $\phi_{\mathrm{ext}} = 0°$ corresponds to a shearing along the θ_1 direction whereas $\phi_{\mathrm{ext}} = 90°$ corresponds to a shearing in the θ_2 direction.

In Figure 1.16, we show the critical curves and caustics of an NIS with an external shear of strength $\gamma_{\mathrm{ext}} = 0.2$ and position angle $\theta_{\mathrm{ext}} = 0$. The coordinates have been rescaled by θ_{e}, i.e. $x_1 = \theta_1/\theta_{\mathrm{e}}$, $y_1 = \beta_1/\theta_{\mathrm{e}}$, and similarly for x_2 and y_2. In the scaled coordinates, $x_{\mathrm{c}} \equiv \theta_{\mathrm{c}}/\theta_{\mathrm{e}}$ is the characteristic core size. The solid symbols indicate various source positions, and the open symbols of the same shapes are the corresponding image positions.

For large x_{c}, there is only one critical curve (dotted) and one corresponding caustic (solid) as shown in the top-left panel. Inside the caustic, there are three images of the source, whereas outside the caustic there is only one image of the source. As the characteristic core size of the NIS (x_{c}) decreases, a second set of critical/caustic curves appears, which is illustrated in the remaining panels of the figure. Outside the caustic curves, the source has only one image, and every time the source crosses a caustic curve from the outside, two additional images are created. Note that when the source is located within the two sets of caustic curves, there are five images of the source. The fifth central image exists as a result of the non-singular nature of the lens mass distribution. Nonetheless, this fifth image is usually demagnified compared to the other images.

In the bottom-right panel, the critical curves and caustics are similar to the ones that were discussed in Section 1.3.6 and shown in Figure 1.12. The solid triangle is just within the fold of the astroid, leading to the fold configuration of the set of images marked by open triangles. The solid square is just within the upper cusp of the astroid, corresponding to the cusp configuration of the set of images marked by open squares.

As shown in the bottom-left panel, when the core is relatively large, the cusp of the astroid caustic can extend beyond the radial caustics. In this case, if a source is close to the cusp, as indicated by the solid hexagon, the so-called 'naked cusp' configuration
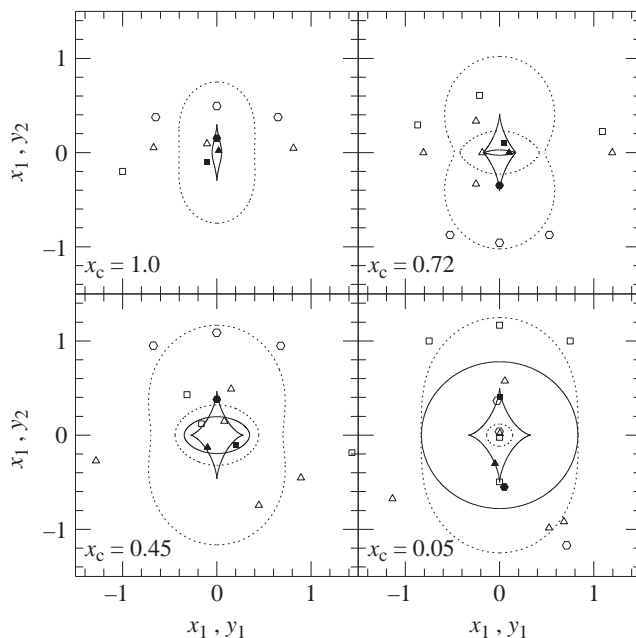
FIGURE 1.16. Non-singular isothermal sphere (NIS) with external shear. Caustics (solid) and critical curves (dashed) are shown for four characteristic core sizes x_c of the NIS which are labelled on the panels. In each panel, various image configurations are illustrated by different symbol shapes. The solid symbols indicate source positions, and the corresponding open symbols mark the image positions. (Reproduced, with the permission of P. Schneider, from Schneider et al. (eds) 2006, *Gravitational Lensing: Strong, Weak and Micro* (Berlin: Springer-Verlag).)

arises where three images (open hexagon symbols) appear on one side of the lens without a central or counter image on the opposite side.

1.4.3.2 NIE

One way to parametrize the non-singular isothermal ellipsoid (NIE) mass distribution is

$$\kappa_{\mathrm{NIE}}(\theta_1, \theta_2) = \frac{\theta_{\mathrm{e}}}{2\sqrt{\theta_1^2 + \theta_2^2/q^2 + b_{\mathrm{c}}^2}}, \tag{1.86}$$

where θ_{e} is the characteristic strength of the NIE, q is the axis ratio of the elliptical isodensity contours and b_{c} is the core radius. For $q = 1$, this distribution reduces to the NIS, and for $b_{\mathrm{c}} \to 0$, the distribution reduces to the SIE (singular isothermal ellipsoid).

The typical critical curves and caustics of the NIE were already shown in Figure 1.12 and its general properties were discussed in Section 1.3.6. More details on the NIE can be found in, for example, Kassiola & Kovner (1993) and Kormann, Schneider and Bartelmann (1994).

1.4.4 *Power-law lens mass distributions and their applications*

What are the mass density profiles of the galaxies in the Universe? The isothermal profiles that we considered in the previous sections produce flat rotation curves somewhat similar to those observed in many massive galaxies, but are the density profiles of galaxies really isothermal? To test the assumption of isothermality, one can model the lens mass density profile by a power-law $\rho(r) \propto r^{-\gamma'}$ and measure the power-law slope γ'. The case of $\gamma' = 2$

corresponds to isothermality. The lensing properties, such as the deflection angles, of the power-law profile can be computed quickly following Barkana (1998).

Using observations of strong lensing and lens stellar velocity dispersions of a sample of strong lens galaxies, the SLACS (Sloan Lens ACS survey) team have constrained the power-law slopes of the lens galaxies and found that on average the galaxies are very close to isothermal with a mean $\gamma' \sim 2.08$ and a $\sim 10\%$ intrinsic scatter on the slope (e.g. Koopmans et al. 2009; Auger et al. 2010; Barnabè et al. 2011). Both the mean slope and the scatter can shed light on the feedback mechanisms of galaxy formation (Dutton & Treu 2014). By analysing the weak lensing signal around 22 SLACS galaxies, Gavazzi el al. (2007) showed that the total mass profile of these galaxies is consistent with being isothermal between 1 and 100 effective radius. By decomposing the total mass distribution into a luminous component and a dark component, various studies of the SLACS lenses (e.g. Grillo et al. 2009; Treu et al. 2010; Auger et al. 2010; Barnabè et al. 2011) found that the Salpeter stellar initial mass function (IMF) fits the observations of the lenses better than Chabrier or Kroupa. Furthermore, more massive systems contain more dark matter, and that dark matter already dominates in the inner regions. This implies a mass-dependent central dark matter distribution and/or a non-universal stellar IMF. Grillo et al. (2009) and Treu et al. (2009) showed that the SLACS lens galaxies are not different from twins of non-lens galaxies, so the results of the SLACS lenses can be generalized to the overall population of early-type galaxies.

The SLACS studies have shown that galaxies at $z \sim 0.2$ are well described by power-law lens mass distributions and have nearly isothermal profiles. What about galaxies at higher redshifts? Two surveys have been conducted to search systematically for higher redshift lenses. SL2S (Strong Lensing Legacy Survey; Gavazzi et al. 2012) looked through the imaging data of the Canada–France–Hawaii Telescope Legacy Survey for lenses, while BELLS (BOSS Emission-Line Lens Survey; Brownstein et al. 2012) followed a similar technique as SLACS and looked for lens candidates in spectra from BOSS (Baryon Oscillation Spectroscopic Survey; Dawson et al. 2013). Both surveys have dozens of lenses at $z \sim 0.5$. By combining with the SLACS lenses at lower redshift, both teams detected evolution in the total mass density profile of lens galaxies in that the profiles got steeper with time (Ruff et al. 2011; Bolton et al. 2012). The surveys are ongoing to place tighter constraints on the evolution of the structures of galaxies.

Power-law mass distribution has been used to model lens galaxies for cosmological studies. As discussed in Section 1.3.2, strong lens systems with time delays measured between the multiple images and with models of the lens mass distribution can be used to determine $D_{\Delta t}$ to the lens and constrain cosmological parameters. By modelling the lens potential on a grid of pixels, Suyu et al. (2009) demonstrated that the lens galaxy profiles can be well described by power-law profiles even in a complicated lens system with lens galaxies that appear to be interacting. The time-delay distance depends sensitively on the slope of the power-law profile: the steeper the profile slope, the lower the time-delay distance for fixed (measured) time delays between the multiple images. The intrinsic scatter in the slope of galaxies from SLACS would affect cosmological inferences at the $\sim 20\%$ level, which is significantly larger than the uncertainties of the best current cosmological probes; therefore, the slopes of the time-delay lenses need to be measured directly and more precisely in order to determine $D_{\Delta t}$ to a few per cent. Multiple compact source components, or spatially extended source surface brightness distributions can be used to constrain the slope of the lens profile at the location of the multiple image, where it matters (e.g. Wucknitz, Biggs and Browne 2004; Dye et al. 2008; Suyu 2012). Recent studies based on time-delay measurements from radio monitoring (e.g. Fassnacht et al. 2002) and optical lens monitoring particularly from COSMOGRAIL

(COSmological MOnitoring of GRAvItational Lenses; e.g. Courbin et al. 2011; Tewes et al. 2013) and the Kochanek et al. (2006) team have yielded time-delay distance measurements for two lenses, each with ∼5–6% uncertainty (Suyu et al. 2010, 2013). The cosmological results based on the two lenses are comparable in precision and complementary to other cosmological probes, including supernovae and baryon acoustic oscillations, when each probe is combined with the results of the cosmic microwave background.

To summarize, despite the simplicity of the form of the power-law profile, it has been successfully applied to real gravitational lens systems to study, for example, galaxy structures and cosmology.

Acknowledgements

REFERENCES

Alcock, C. et al. 1993, *Nature*, **365**, 621

Applegate, D. E. et al. 2014, *MNRAS*, **439**, 48

Aubourg, E. et al. 1993, *Nature*, **365**, 623

Auger, M. W., Treu, T., Bolton, A. S., Gavazzi, R., Koopmans, L. V. E., Marshall, P. J., Moustakas, L. A. & Burles, S. 2010, *ApJ*, **724**, 511

Bachelet, E. et al. 2012, *ApJ*, **754**, 73

Barkana, R. 1998, *ApJ*, **502**, 531

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., Bolton, A. S. & Gavazzi, R. 2009, *MNRAS*, **399**, 21

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T. & Bolton, A. S. 2011, *MNRAS*, **415**, 2215

Bartelmann, M. 2010, *Classical and Quantum Gravity*, **27**, 233001

Beaulieu, J.-P. et al. 2006, *Nature*, **439**, 437

Benson, B. A. et al. 2013, *ApJ*, **763**, 147

Blandford, R. & Narayan, R. 1986, *ApJ*, **310**, 568

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T. & Moustakas, L. A. 2006, *ApJ*, **638**, 703

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., Gavazzi, R., Moustakas, L. A., Wayth, R. & Schlegel, D. J. 2008, *ApJ*, **682**, 964

Bolton, A. S. et al. 2012, *ApJ*, **757**, 82

Bradač, M. et al. 2009, *ApJ*, **706**, 1201

Brainerd, T. G., Blandford, R. D. & Smail, I. 1996, *ApJ*, **466**, 623

Browne, I. W. A. et al. 2003, *MNRAS*, **341**, 13

Brownstein, J. R. et al. 2012, *ApJ*, **744**, 41

Burke, W. L. 1981, *ApJL*, **244**, L1

Coe, D. et al. 2013, *ApJ*, **762**, 32

Courbin, F. et al. 2011, *A&A*, **536**, A53

Dalal, N. & Kochanek, C. S. 2002, *ApJ*, **572**, 25

Dawson, K. S. et al. 2013, *AJ*, **145**, 10

Dutton, A. A. & Treu, T. 2014, *MNRAS*, **438**, 3594

Dye, S., Evans, N. W., Belokurov, V., Warren, S. J. & Hewett, P. 2008, *MNRAS*, **388**, 384

Fadely, R. & Keeton, C. R. 2012, *MNRAS*, **419**, 936

Fadely, R., Keeton, C. R., Nakajima, R. & Bernstein, G. M. 2010, *ApJ*, **711**, 246

Falco, E. E., Gorenstein, M. V. & Shapiro, I. I. 1985, *ApJL*, **289**, L1

Fassnacht, C. D., Xanthopoulos, E., Koopmans, L. V. E. & Rusin, D. 2002, *ApJ*, **581**, 823

Fassnacht, C. D., Gal, R. R., Lubin, L. M., McKean, J. P., Squires, G. K. & Readhead, A. C. S. 2006, *ApJ*, **642**, 30

Fassnacht, C. D., Koopmans, L. V. E. & Wong, K. C. 2011, *MNRAS*, **410**, 2167

Faure, C. et al. 2008, *ApJS*, **176**, 19

Gavazzi, R., Treu, T., Rhodes, J. D., Koopmans, L. V. E., Bolton, A. S., Burles, S., Massey, R. J. & Moustakas, L. A. 2007, *ApJ*, **667**, 176

Gavazzi, R., Treu, T., Marshall, P. J., Brault, F. & Ruff, A. 2012, *ApJ*, **761**, 170

Gorenstein, M. V., Shapiro, I. I. & Falco, E. E. 1988, *ApJ*, **327**, 693

Grillo, C., Gobat, R., Lombardi, M. & Rosati, P. 2009, *ApJ*, **501**, 461

Hezaveh, Y. D. et al. 2013, *ApJ*, **767**, 132

Inada, N. et al. 2012, *AJ*, **143**, 119

Irwin, M. J., Webster, R. L., Hewett, P. C., Corrigan, R. T. & Jedrzejewski, R. I. 1989, *AJ*, **98**, 1989

Jackson, N. 2008, *MNRAS*, **389**, 1311

Jackson, N., Rampadarath, H., Ofek, E. O., Oguri, M. & Shin, M.-S. 2012, *MNRAS*, **419**, 2014

Jullo, E., Natarajan, P., Kneib, J.-P., D'Aloisio, A., Limousin, M., Richard, J. & Schmid, C. 2010, *Science*, **329**, 924

Kassiola, A. & Kovner, I. 1993, *ApJ*, **417**, 450

Keeton, C. R. & Zabludoff, A. I. 2004, *ApJ*, **612**, 660

Kilbinger M. et al. 2013, *MNRAS*, **430**, 2200

Kochanek, C. S., Morgan, N. D., Falco, E. E., McLeod, B. A., Winn, J. N., Dembicky, J. & Ketzeback, B. 2006, *ApJ*, **640**, 47

Koopmans, L. V. E. & Treu, T. 2003, *ApJ*, **583**, 606

Koopmans, L. V. E. et al. 2009, *ApJL*, **703**, L51

Kormann, R., Schneider, P. & Bartelmann, M. 1994, *A&A*, **284**, 285

Limousin, M. et al. 2009, *A&A*, **502**, 445

Limousin, M., Morandi, A., Sereno, M., Meneghetti, M., Ettori, S., Bartelmann, M. & Verdugo, T. 2013, *Space Science Reviews*, **177**, 155

Lynds, R. & Petrosian, V. 1986, *BAAS*, **18**, 1014

Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M. & Brinkmann, J. 2006, *MNRAS*, **368**, 715

Marshall, P. J. et al. 2007, *ApJ*, **671**, 1196

Marshall, P. J., Hogg, D. W., Moustakas, L. A., Fassnacht, C. D., Bradač, M., Schrabback, T. & Blandford, R. D. 2009, *ApJ*, **694**, 924

McKean, J. P., Koopmans, L. V. E., Flack, C. E., Fassnacht, C. D., Thompson, D., Matthews, K., Blandford, R. D., Readhead, A. C. S. & Soifer, B. T. 2007, *MNRAS*, **378**, 109

Momcheva, I., Williams, K., Keeton, C. & Zabludoff, A. 2006, *ApJ*, **641**, 169

Morgan, C. W., Kochanek, C. S., Morgan, N. D. & Falco, E. E. 2006, *ApJ*, **647**, 874

Morgan, C. W., Eyler, M. E., Kochanek, C. S., Morgan, N. D., Falco, E. E., Vuissoz, C., Courbin, F. & Meylan, G. 2008, *ApJ*, **676**, 80

Mosquera, A. M., Kochanek, C. S., Chen, B., Dai, X., Blackburne, J. A. & Chartas, G. 2013, *ApJ*, **769**, 53

Myers, S. T. et al. 2003, *MNRAS*, **341**, 1

Natarajan, P., Kneib, J.-P., Smail, I., Treu, T., Ellis, R., Moran, S., Limousin, M. & Czoske, O. 2009, *ApJ*, **693**, 970

Negrello, M. et al. 2010, *Science*, **330**, 800

Newman, A. B., Treu, T., Ellis, R. S. & Sand, D. J. 2013, *ApJ*, **765**, 25

Oguri, M. et al. 2006, *AJ*, **132**, 999

Okabe, N., Smith, G. P., Umetsu, K., Takada, M. & Futamase, T. 2013, *ApJL*, **769**, L35

Paczyński, B. 1986, *ApJ*, **304**, 1

Paczyński, B. 1987, *Nature*, **325**, 572

Refsdal, S. 1964, *MNRAS*, **128**, 307

Refsdal, S. 1966, *MNRAS*, **132**, 101

Ruff, A. J., Gavazzi, R., Marshall, P. J., Treu, T., Auger, M. W. & Brault, F. 2011, *ApJ*, **727**, 96

Schneider, P. 1984, *A&A*, **140**, 119

Schneider, P. 1985, *A&A*, **143**, 413

Schneider, P., Ehlers, J. & Falco, E. E. 1992, *Gravitational Lenses: XIV* (Berlin: Springer-Verlag)

Schneider, P., Kochanek, C. S. & Wambsganss, J. 2006, *Gravitational Lensing: Strong, Weak and Micro* (Berlin: Springer-Verlag)

Schrabback, T. et al. 2010, *A&A*, **516**, A63

Sonnenfeld, A., Treu, T., Gavazzi, R., Marshall, P. J., Auger, M. W., Suyu, S. H., Koopmans, L. V. E. & Bolton, A. S. 2012, *ApJ*, **752**, 163

Soucail, G., Fort, B., Mellier, Y. & Picat, J. P. 1987, *A&A*, **172**, L14

Soucail, G., Mellier, Y., Fort, B., Mathez, G. & Cailloux, M. 1988, *A&A*, **191**, L19

Suyu, S. H. 2012, *MNRAS*, **426**, 868

Suyu, S. H. & Halkola, A. 2010, *A&A*, **524**, A94

Suyu, S. H., Marshall, P. J., Blandford, R. D., Fassnacht, C. D., Koopmans, L. V. E., McKean, J. P. & Treu, T. 2009, *ApJ*, **691**, 277

Suyu, S. H., Marshall, P. J., Auger, M. W., Hilbert, S., Blandford, R. D., Koopmans, L. V. E., Fassnacht, C. D. & Treu, T. 2010, *ApJ*, **711**, 201

Suyu, S. H. et al. 2013, *ApJ*, **766**, 70

Tewes, M. et al. 2013, *A&A*, **556**, A22

Treu, T. 2010, *ARA&A*, **48**, 87

Treu, T. & Koopmans, L. V. E. 2002, *MNRAS*, **337**, L6

Treu, T. & Koopmans, L. V. E. 2004, *ApJ*, **611**, 739

Treu, T., Gavazzi, R., Gorecki, A., Marshall, P. J., Koopmans, L. V. E., Bolton, A. S., Moustakas, L. A. & Burles, S. 2009, *ApJ*, **690**, 670

Treu, T., Auger, M. W., Koopmans, L. V. E., Gavazzi, R., Marshall, P. J. & Bolton, A. S. 2010, *ApJ*, **709**, 1195

Treu, T., Dutton, A. A., Auger, M. W., Marshall, P. J., Bolton, A. S., Brewer, B. J., Koo, D. C. & Koopmans, L. V. E. 2011, *MNRAS*, **417**, 1601

Tyson, J. A., Wenk, R. A. & Valdes, F. 1990, *ApJL*, **349**, L1

Vegetti, S., Lagattuta, D. J., McKean, J. P., Auger, M. W., Fassnacht, C. D. & Koopmans, L. V. E. 2012, *Nature*, **481**, 341

Vieira, J. D. et al. 2013, *Nature*, **495**, 344

von der Linden, A. et al. 2014, *MNRAS*, **439**, 2

Walsh, D., Carswell, R. F. & Weymann, R. J. 1979, *Nature*, **279**, 381

Winn, J. N., Rusin, D. & Kochanek, C. S. 2004, *Nature*, **427**, 613

Wong, K. C., Keeton, C. R., Williams, K. A., Momcheva, I. G. & Zabludoff, A. I. 2011, *ApJ*, **726**, 84

Wucknitz, O., Biggs, A. D. & Browne, I. W. A. 2004, *MNRAS*, **349**, 14

Young, P., Gunn, J. E., Oke, J. B., Westphal, J. A. & Kristian, J. 1980, *ApJ*, **241**, 507